

# What Snippets Say About Pages (Abstract)

T. Demeester  
Ghent University  
tdmeeste@ugent.be

D. Nguyen  
University of Twente  
d.nguyen@utwente.nl

D. Trieschnigg  
University of Twente  
d.trieschnigg@utwente.nl

C. Develder  
Ghent University  
cdvelder@ugent.be

D. Hiemstra  
University of Twente  
d.hiemstra@utwente.nl

## ABSTRACT

We summarize findings from [1]. What is the likelihood that a Web page is considered relevant to a query, given the relevance assessment of the corresponding snippet? Using a new Federated Web Search test collection that contains search results from over a hundred search engines on the internet, we are able to investigate such research questions from a global perspective. Our test collection covers the main Web search engines like Google, Yahoo!, and Bing, as well as smaller search engines dedicated to multimedia, shopping, etc., and as such reflects a realistic Web environment. Using a large set of relevance assessments, we are able to investigate the connection between snippet quality and page relevance. The dataset is strongly heterogeneous, and care is required when comparing resources. To this end, a number of probabilistic variables, based on snippet and page relevance, are introduced and discussed.

## 1. INTRODUCTION

Finding our way around among the vast quantities of data on the Web would be unthinkable without the use of Web search engines. Apart from a limited number of very large search engines that constantly crawl the Web for publicly available data, a large amount of smaller and more focused search engines exist, specialized in specific information goals or data types (e.g., online shopping, news, multimedia, social media). In order to promote research on Federated Web Search, we created a large dataset containing sampled results from 108 search engines on the internet, and containing relevance judgments for the top 10 results (both snippets and pages) from all of these resources for 50 test topics (from the TREC 2010 Web Track). The relevance judgments are particularly interesting for analysis, partly because they originate from very diverse collections (both in size and in scope, whereby the relevance judgments are done in a generic way), and partly because we not only judged the result pages, but also, independently, the original snippets. Our analysis deals with ranked result lists from diverse retrieval algorithms, and with snippets from various snippet generation strategies, as they are currently in use on the Web.

This abstract is based on [1], which has the following scope.

*DIR 2013*, April 26, 2013, Delft, The Netherlands.

First, after an overview of related work, the relevance judgments for the new dataset are discussed at length, with emphasis on the assessors' consistency. Second, a number of potential difficulties in Federated Web Search and especially in the evaluation of relevance are discussed, related to the heterogeneous character of the resources. Finally, a probabilistic analysis of the relationship between the indicative snippet relevance and the actual page relevance is presented (where by 'page' we denote a result item like a web page, a video, scientific paper... as returned by the included search engines). In a further contribution [2], it is shown that the information carried by an average snippet can be used to make a reasonable prediction of the relevance of the result page itself. Within the limits of this abstract, we will primarily focus on the question of why the user's snippet-based prior estimation of the page relevance is of paramount importance for the overall performance of the search service. Using the relevance judgments for the dataset presented in [3], the relevant concepts are illustrated for the specific case of large general web search engines.

## 2. SNIPPET VS. PAGE RELEVANCE

The intuition behind this paper is simple: a search engine can only exploit the full potential of its retrieval algorithm if the result snippets reflect the relevance of the corresponding pages as well as possible. This means that a highly relevant result should be presented to the user by a very promising snippet, and a less relevant result page by a less interesting snippet. If there is a mismatch between what the user estimates from a result snippet and the actual result page, the overall performance of the system degrades.

For a more formal analysis, we introduce the snippet relevance variable  $S$ , and the page relevance variable  $P$ . As for the specific relevance levels, the snippet relevance  $S$  ranges from **No**, over **Unlikely** and **Maybe**, to **Sure**, indicating how likely the assessor estimates the result page behind the snippet to be relevant. The levels for  $P$ , the page relevance, are **Non**, **Rel** (containing minimal relevant information), **HRel** (highly relevant), **Key** (worthy of being a top result), and **Nav** (for navigational queries). In this paper we will either indicate the considered relevance level explicitly, such as  $S = \text{Sure}$  (i.e., considering only snippets with the label **Sure**), or define binary relevance levels, such as  $P \geq \text{HRel}$  (indicating page relevance levels of **HRel**, **Key**, or **Nav**).

Table 1: Overview of the relationship between page and snippet judgments, for different types of resources, and based on the page relevance level  $P \geq \text{HRel}$ .

	S=Unlikely	S=Maybe	S=Sure		
	$\mathcal{P}(P S)$	$\mathcal{P}(P S)$	$\mathcal{P}(P S)$	$\mathcal{P}(P,S)$	$\mathcal{P}(P)$
General Web search	0.20	0.40	0.65	0.26	0.34
Multimedia	0.09	0.23	0.48	0.06	0.09
News	0.09	0.19	0.42	0.02	0.03
Shopping	0.06	0.10	0.21	0.01	0.03
Encyclopedia/Dict	0.05	0.23	0.58	0.11	0.14
Books	0.12	0.10	0.18	0.02	0.05
Blogs	0.12	0.23	0.40	0.05	0.07

Table 2: Comparison of the largest general Web search engines

	$\mathcal{P}(S=\text{Sure})$	$P \geq \text{HRel}$ and $S=\text{Sure}$			$P \geq \text{Key}$ and $S=\text{Sure}$		
		$\mathcal{P}(P S)$	$\mathcal{P}(P,S)$	$\mathcal{P}(P)$	$\mathcal{P}(P S)$	$\mathcal{P}(P,S)$	$\mathcal{P}(P)$
Google	0.42	0.68	0.28	0.38	0.39	0.16	0.19
Yahoo!	<b>0.47</b>	0.69	<b>0.32</b>	<b>0.44</b>	0.38	0.18	<b>0.22</b>
Bing	0.41	0.60	0.24	0.28	0.30	0.12	0.13
Baidu	0.21	0.43	0.09	0.12	0.23	0.05	0.06
Mamma.com	0.43	<b>0.73</b>	0.31	0.41	<b>0.44</b>	<b>0.19</b>	<b>0.22</b>

Retrieval systems are typically being evaluated based on the probability of relevance of the result page, written  $\mathcal{P}(P)$ . If however the access to that page also depends on the user’s estimate of a snippet, the actual measure to consider should be  $\mathcal{P}(P,S)$ , the mutual probability of relevance for both the snippet and the page. Note that it can be written as  $\mathcal{P}(S)\mathcal{P}(P|S)$ , in which  $\mathcal{P}(P|S)$  is the conditional probability of the page label, given the snippet label. Studying  $\mathcal{P}(P|S)$  is especially instructive, for instance to find out how often a relevant page remains hidden behind a non-convincing snippet.

For several resource categories, table 1 gives empirical estimates of such probabilities for binary page relevance  $P \geq \text{HRel}$ , based on our relevance judgements. Comparing  $\mathcal{P}(P|S)$  for the snippet labels *Maybe* and *Sure* shows that a relatively large amount of *HRel* pages are behind snippets which were judged only *Maybe*, especially for the general search engines. This shows that often a *HRel* page’s snippet cannot convince the user that the page is indeed highly relevant. We also observe that for the snippet label  $S=\text{Sure}$ , e.g., the News resources display a relatively high  $\mathcal{P}(P|S)$ , against a very low  $\mathcal{P}(P,S)$ . In other words, these resources returned only very few relevant results for our test topics, but if a snippet was found relevant, 4 out of 10 times it points to one of those few relevant results.

As the test topics are best suited for the general Web search engines, we can explicitly compare the performance of four of the largest general Web search engines in our collection, i.e., Google, Yahoo!, Bing, and Baidu, as well as Mamma.com, which is actually a metasearch engine. Table 2 presents the results. It appears that for the snippet label  $S=\text{Sure}$  and two page relevance levels ( $P \geq \text{HRel}$  and  $P \geq \text{Key}$ ),  $\mathcal{P}(P,S)$  is consistently lower than  $\mathcal{P}(P)$ , which is actually the averaged precision@10 of page relevance, and does not take into account the fact that the snippet is not always as promising as the page is relevant. The metasearch engine outperforms the others, as it aggregates results from a number of resources, such as Google, Yahoo!, and Bing. We want to stress that the considered test topics are still no representative collec-

tion of, for example, popular Web queries, and therefore we cannot draw any further conclusions about these search engines beyond the scope of our test collection. Yet, here is another example of how the table might be interpreted, with that in mind. Considering only *Key* results, we could compare Yahoo! and Bing. Yahoo! seems to score higher for all reported parameters, so either Bing’s collection contains a smaller number of relevant results, or Yahoo!’s retrieval algorithms are better tuned for our topics. The lower value of  $\mathcal{P}(P|S)$  for Bing shows that it has a slightly increased chance that the page for a promising snippet appears less relevant. However, the ratio of  $\mathcal{P}(P,S)$  and  $\mathcal{P}(P)$  is higher for Bing than for Yahoo!, indicating that for Yahoo!, its own recall on *Key* pages will be decreased more due to the quality of the snippets, than for Bing. In fact, we found that  $\mathcal{P}(S=\text{Sure}|P \geq \text{Key})$  is 79% for Yahoo!, but 91% for Bing.

### 3. CONCLUSIONS

Analyzing the relationship between the relevance of snippets from a large amount of on-line search engines and the relevance of the corresponding result pages, clearly shows that in the evaluation of and comparison between different resources, the snippets cannot be left out.

### 4. ACKNOWLEDGMENTS

This research was partly supported by the Netherlands Organization for Scientific Research, NWO, grants 639.022.809 and 640.005.002, and partly by iMinds in Flanders.

### 5. REFERENCES

- [1] T. Demeester, D. Nguyen, D. Trieschnigg, C. Develder, and D. Hiemstra. What Snippets Say about Pages in Federated Web Search. In *AIRS*, 2012.
- [2] T. Demeester, D. Nguyen, D. Trieschnigg, C. Develder, and D. Hiemstra. Snippet-Based Relevance Predictions for Federated Web Search. In *ECIR*, 2013.
- [3] D. Nguyen, T. Demeester, D. Trieschnigg, and D. Hiemstra. Federated Search in the Wild: the Combined Power of over a Hundred Search Engines. In *CIKM*, 2012.